Manfred M. Fischer and James P. LeSage

A Bayesian approach to identifying and interpreting regional convergence clubs in Europe

Paper

A Bayesian approach to identifying and interpreting regional convergence clubs in Europe

Manfred M. Fischer
Department for SocioEconomics
Vienna University of Economics and Business
Vienna, Austria
E-mail: Manfred.Fischer@wu.ac.at
and
James P. LeSage
Fields Endowed Chair
Texas State University-San Marcos
Department of Finance & Economics
San Marcos, TX 78666
E-mail: jlesage@spatial-econometrics.com

October, 2012

Abstract

This study suggests a two-step approach to identifying and interpreting regional convergence clubs in Europe. The first step involves identifying the number and composition of clubs using a space-time panel data model for annual income growth rates in conjunction with Bayesian model comparison methods. A second step uses a Bayesian space-time panel data model to assess how changes in the initial endowments of variables (that explain growth) impact regional income levels over time. These dynamic trajectories of changes in regional income levels over time allow us to draw inferences regarding the timing and magnitude of regional income responses to changes in the initial conditions for the clubs that have been identified in the first step. This is in contrast to conventional practice that involves setting the number of clubs ex ante, selecting the composition of the potential convergence clubs according to some a priori criterion (such as initial per capita income thresholds for example), and using cross-sectional growth regressions for estimation and interpretation purposes.

KEYWORDS: Dynamic space-time panel data model, Bayesian Model Comparison, European regions.

JEL: C11, C23, O47, O52

1 Introduction

The question whether incomes are converging across regions has attracted the attention of both growth economists and regional scientists (see Durlauf, Johnson and Temple 2005; and Magrini 2004 for useful surveys). The bulk of the empirical literature on this question has focused on growth regressions of the type pioneered by Barro (1991), and Mankiw, Romer and Weil (1992). Recent work has extended growth analysis to consider panels (see, for example, Islam 1995; Lee, Pesaran and Smith 1997) and/or to account for spatial effects among regions (see, for example, Fingleton and López-Bazo 2006; and LeSage and Fischer 2008; and for theoretical underpinnings Ertur and Koch 2007; and Fischer 2011). In spite of the large work done, relatively little explicit attention has been paid to the question of systematically identifying and interpreting convergence clubs.

The notion of club convergence can be traced back to Baumol (1986), but owes its more rigorous formulation to Durlauf and Johnson (1995), and Galor (1996). The concept is based on new growth theories that yield multiple, locally stable steady state equilibria in per capita output.[1] In contrast to conventional wisdom Galor (1996) has demonstrated if heterogeneity is permitted across individuals, multiplicity of stationary equlibria may also occur in Solow and Mankiw-Romer-Weil worlds, and in these cases the distribution of initial income per capita determines the club to which a particular region will belong.[2] But neither neoclassical nor new growth theories offer explicit guidance in determining the number and composition of clubs within a given cross-section of regions.

The standard approach to this problem in club convergence analysis involves setting the number of clubs ex ante, selecting the composition of the potential convergence clubs according to some a priori criterion (such as initial income per capita thresholds), and then using cross-sectional growth regressions for estimation and interpretation purposes. Examples include Dall'erba (2005); Mora (2005); Ertur, LeGallo and Baumont (2006); and Fischer

[1]Modern growth theory has suggested that the distribution of per capita income of regions may display a tendency for the steady state distribution to cluster around a small number of poles of attraction, and hence lead to convergence clubs (Canova 2004). This tendency may be due to several factors: capital market imperfections, externalities, non-convexities, and imperfectly competitive market structures (Galor 1996).

[2]Regions that are similar in their structural characteristics, but differ in their initial distribution of income, may cluster around different steady state equilibria (see Durlauf 1996; Quah 1996). It should be noted that if multiple equlibria depend on initial income cut-offs, the relationship between subsequent growth and initial income will *not* be linear.

and Stirböck (2006). Dissatisfaction with this approach has generated an increasingly large amount of literature, employing a wide variety of statistical methods.[3]

An early effort to this line of research goes back to Durlauf and Johnson (1995) who use classification and regression tree methods to search for non-linearities in the growth process as implied by the existence of convergence clubs.[4] Another important, but more recent approach is due to Canova (2004) who introduces a procedure for panel data that establishes the number of groups (clubs) and the assignment of regions to these clubs, drawing on Bayesian ideas to test for unknown break-points in the time series. In contrast to Durlauf and Johnson (1995), this approach shows the important feature that it allows for parameter heterogeneity across regions within a club. Heterogeneity takes the form of a prior that restricts the coefficients of the regions in a club to have the same distribution, but allows the distribution of the coefficients of regions in different clubs to differ. The approach allows to order the regions by various criteria (such as, for example, initial per capita income). The estimation procedure then selects break points and group membership by maximizing the predictive density (marginal likelihood) of the data with respect to the location of the break points and group membership.

The objective of our study is to develop a novel approach to identify the number and composition of convergence clubs within a given cross-section of European regions. The study lies in the research tradition that finds it useful to view multiple growth regimes as evidence for the existence of convergence clubs.[5] Our work is related to the study by Canova (2004) in so far that we also draw on Bayesian ideas to identify regional convergence clubs in Europe.

The analysis, however, differs from this and other previous research in at least two major respects. *First*, we attempt to identify sets of regions (clubs) that obey separate

[3]Hobijn and Franses (2000), for example, suggest using a cluster algorithm to endogenously identify groups of converging countries or regions. But in the absence of controls for structural characteristics it is not clear whether these clusters represent groups of countries or regions in distinct basins of attraction of the growth process. Corrado, Martin and Weeks (2005) extend this approach to allow for time variation in clusters. Desdoigts (1999) makes use of projection pursuit methods in an attempt to identify groups of countries with relatively homogenous growth experiences based on data about the characteristics and initial conditions of each country. Phillips and Sul (2009), utilize a clustering mechanism test procedure that relies on a stepwise and cross-section recursive application of log t regression tests.

[4]See De Siana and D'Uva (2006) for a more recent application of this approach to European regions.

[5]But it is not clear whether they represent groups of regions in distinct basins of attraction of the growth process. This so-called identification problem is outside the scope of this paper.

growth regimes with regime membership determined using Bayesian dynamic space-time panel data comparison methods. *Second*, we employ a Bayesian dynamic space-time panel data model to estimate the parameters for each club suggested by the Bayesian classification scheme. We derive analytical expressions for the partial derivative impacts of changes in the initial endowments on regional levels of income over time. These regional trajectories allow inferences regarding the timing and magnitude of (direct and indirect) regional income response elasticities to changes in the initial conditions for the clubs, and these trajectories provide clear evidence of the distinct long-term behaviour of the clubs.

The rest of the paper is organized as follows. Section 2 outlines the dynamic space-time panel data model applied to annual (per capita) income growth rates, and the formal Bayesian model comparison methodology as it applies to our work here.[6] A key insight is that each assignment of regions to a particular club membership can be viewed as a distinct model. This allows formal model comparison methods to use, so the model (sample split) with the highest posterior model probability for a given number of clubs can be established. Of course, the resulting club classification is conditional on the dynamic space-time panel income growth rates model specification used in the comparison procedure. The empirically determined club assignments are reported in Section 3.

Section 4 describes the second step of our approach, which uses a dynamic space-time panel data model to analyze the space-time dynamic relationship between regional levels of income over time and space.[7] The model includes spatial and temporal dependence as well as space-time covariance so that changes in the endowments of a single own-region (say i) at time t can impact own- and other-regions ($j \neq i$) in the current and future time periods. In particular, we focus on the partial derivative impact of changes in the regional endowment variables in the matrix X_t on regional income levels Y_{t+T} at various time horizons T, an issue that has received little attention in the spatial panel data model literature.[8] The final section summarizes and concludes.

[6] Of course, there is a relationship between growth rates and level values taken by variables (such as income, physical and human capital) over time which is explored in detail for the case of spatially dependence sample data in LeSage and Fischer (2008).

[7] The motivation for the use of this model type is that it can provide us with useful information about the clubs of regions not available from cross-section (spatial) regressions.

[8] Parent and LeSage (2010) as well as Debarsy, Ertur and LeSage (2012) are exceptions.

2 The methodology for identifying clubs

The first step of our approach uses a formal Bayesian model comparison methodology to classify European regions into clubs. Each region must be classified into one of M clubs. The classification takes place conditional on a space-time (random effects) panel data model[9] of regional income growth given by

$$\begin{aligned} g_t &= \phi g_{t-1} + \rho W g_t + \theta W g_{t-1} + \psi \ln y_{t-1} + \alpha \iota_N + X_{t-1}\beta + \eta_t, \qquad (1) \\ \eta_t &= \mu + \varepsilon_t, \\ g_t &= \ln(y_t - y_{t-1}), \qquad t = 2, \dots, T \end{aligned}$$

The panel data growth regression model relates the $N \times 1$ vector of time t growth rates (g_t) to that of the previous time period (g_{t-1}), neighbouring regions in the current time period (Wg_t), and also to that of neighbouring regions in the previous time period (Wg_{t-1}). $g_t = (g_{1t}, \dots, g_{Nt})'$ is the $N \times 1$ vector of observed income growth rates for the tth time period, with y_t denoting income levels at time t, and ψ the parameter reflecting dependence on previous period levels. The intercept parameter is α and ι_N is an $N \times 1$ column vector of ones. Previous period endowments of physical capital stocks, knowledge stocks and human capital which are thought to exert an influence on regional income growth are contained in the $N \times K$ matrix X_{t-1} with K denoting the number of (conditioning) variables included to capture proximate determinants of economic growth and β representing the associated parameter vector.

The vector $\eta_t = \mu + \varepsilon_t$ represents the summation of two unobserved normally distributed random components: μ an $N \times 1$ column vector of random effects with $\mu_i \sim N(0, \sigma_\mu^2), i = 1, \dots, N$, that are fixed across all time periods, and the $N \times 1$ stochastic disturbance ε_t, assumed to be independent and identically distributed with zero mean and scalar variance $\sigma_\varepsilon^2 I_N, t = 1, \dots, T$. We make the traditional assumption that μ is uncorrelated with ε_t for

[9] This type of space-time panel data model specification has been originally proposed by Anselin (2001), and explored by Yu, de Jong and Lee (2008) as well as Parent and LeSage (2011). Examples of empirical studies using this type of specification include Parent and LeSage (2010), and Debarsy, Ertur and LeSage (2012).

identification purposes. W is a known $N \times N$ spatial weight matrix whose diagonal elements are zero. This matrix defines the dependence between cross-sectional (spatial) observational units. We will also assume that W is row-normalized from a symmetric matrix, so that all eigenvalues are real and less than or equal to one. The strength of the spatial dependence is measured by the parameter ρ, the first order time dependence reflected in the scalar parameter ϕ, and θ represents the component mixing space and time dependence.

Our methodology involves use of the panel data variant of traditional cross-sectional growth regressions to identify/classify regions into low- and high-income clubs. The relationship of focus in this model is that between *income growth rates* and the previous period (logged) level of income and (logged) previous period endowments of physical, knowledge and human capital. Given this classification of regions, we use a panel data model (in Section 4.1) that relates (logged) *income levels* to previous period (logged) levels of endowments of physical, knowledge and human capital. This (logged) levels model produces elasticity responses of income levels to endowments that reflect space-time dynamic impacts (marginal effects estimates). The focus is on differences in the space-time dynamic elasticity impacts of endowments on income levels for regions in the two clubs.

Islam (1995) was one of the first studies to examine conventional cross-sectional growth regressions using a panel data setting. He proposed splitting the overall sample into several shorter time spans, with the motivation being that annual growth rates are "too short to be appropriate for studying growth convergence" since "short-term disturbances may loom large in such brief time spans". He relied on five-year time intervals for his panel data model estimation. Despite this brief and informal argument against using annual growth rates in a panel data setting, almost all panel data growth convergence studies have followed the approach of Islam (1995). This includes the space-time dynamic panel data model specification used here in a growth convergence study by Yu and Lee (2012), where four and five year intervals are used.

Our limited sample size of 11 years does not allow us to fully explore this issue. In contrast to growth convergence studies, we use the annual growth rates relationship from (1) in our space-time panel data setting only for the purpose of classifying regions into two clubs. In Section 4.2, our focus is on dynamic elasticity responses of (logged) income levels to previous period (logged) levels of initial period endowments of capital stocks (physical, knowledge and human) using annual time intervals in our panel data model.

In the context of Islam's non-dynamic panel data model, use of initial period endowments from 4 or 5 years ago may make sense. It is less clear how one should proceed for the case of a space-time dynamic model of the type in (1). One possibility is to impose a 3-year lag on the initial period endowment variables as shown in (2). This specification could be justified on the basis for allowing a longer lag between endowment levels from the initial period to influence income growth rates. We present classification results for regions into the two clubs based on models (1) and (2), and these produced relatively similar results.

$$\begin{aligned} g_t &= \phi g_{t-1} + \rho W g_t + \theta W g_{t-1} + \psi \ln y_{t-3} + \alpha \iota_N + X_{t-3}\beta + \eta_t, \qquad (2) \\ \eta_t &= \mu + \varepsilon_t, \\ g_t &= \ln(y_t - y_{t-1}), \qquad t = 4, \ldots, T \end{aligned}$$

The dynamic space-time panel data model relationship in (1) expressed in matrix/vector form shown in (3) is used in conjunction with Bayesian model comparison methods to assign regions to clubs.

$$Pg = H\psi + \iota_{N(T-1)}\alpha + X\beta + \eta \qquad (3)$$

$$P = \begin{pmatrix} B & 0_{N\times N} & 0_{N\times N} & \ldots & 0_{N\times N} \\ A & B & 0_{N\times N} & \ldots & 0_{N\times N} \\ 0_{N\times N} & A & B & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0_{N\times N} \\ 0_{N\times N} & \ldots & 0_{N\times N} & A & B \end{pmatrix} \qquad (4)$$

$$\begin{aligned} H &= \begin{pmatrix} \ln(y_1) & \ldots & \ln(y_{T-1}) \end{pmatrix}' \\ X &= \begin{pmatrix} X_1 & \ldots & X_{T-1} \end{pmatrix}' \\ A &= -(\phi I_N + \theta W) \end{aligned}$$

$$
\begin{aligned}
B &= I_N - \rho W \\
\eta &\sim N(0, \Omega) \\
\Omega &= [(T-1)\sigma_\mu^2 + \sigma_\varepsilon^2](\bar{J}_{T-1} \otimes I_N) + \sigma_\varepsilon^2 \left[(I_{T-1} - \bar{J}_{T-1}) \otimes I_N\right] \qquad (5)
\end{aligned}
$$

We use $\otimes$ to denote the Kronecker product in the expression for Ω in (5), which represents a decomposition proposed by Wansbeek and Kapteyn (1982), that replaces $J_{T-1} = \iota_{T-1}\iota'_{T-1}$ by its idempotent counterpart $\bar{J}_{T-1} = J_{T-1}/(T-1)$ (see Parent and LeSage, 2011).

The scalars σ_μ^2 and σ_ε^2 denote the variances of the random effects vector μ and noise vector ε, respectively. This specification uses the first time period to "feed the lag", leading to the $N(T-1) \times NT$ matrix P in (4). Treating the first period in this way simplifies work involved in analytically calculating the log-marginal likelihood needed to compute posterior probabilities for model comparison purposes, and should have little impact in cases where T is reasonably large.

In our empirical application $N = 216$ European Union regions and $T = 11$ years covering the period from 1995 to 2005, with the initial period being 1995, so T is not excessively large here. To assign regions to candidate clubs we introduce a dummy variable that splits the sample according to initial year (1995) regional income levels above and below m during the initial year 1995. Regions with incomes below m are assigned to Club 1 and those with incomes above this level to Club 2. In (6), we express the dynamic panel model including the $N \times 1$ dummy vector D with zero values for regions where $y_1 \leq m$ and ones for $y_1 > m$, and an $N \times K$ dummy matrix $\tilde{D} = \begin{pmatrix} D & D & \dots & D \end{pmatrix}$. The Hadamard (element-by-element) product $\odot$ is used in conjunction with the dummy matrix $\tilde{D}$ in (6).

$$
\begin{aligned}
g_t &= \phi g_{t-1} + \rho W g_t + \theta W g_{t-1} + \psi \ln(y_{t-1}) + \tilde{\psi} D \ln(y_{t-1}) \\
&+ \alpha \iota_N + \tilde{\alpha} D \iota_N + X_{t-1}\beta + (\tilde{D} \odot X_{t-1})\tilde{\beta} + \eta_t, \qquad t = 2, \dots, T \qquad (6)
\end{aligned}
$$

Parent and LeSage (2011) show that the log-likelihood for this model (with the random effects vector μ integrated out) can be expressed as in (7). For simplicity we use Z to denote a matrix containing all explanatory variables for each time period, and we define:

$\lambda = \sigma_\mu^2/\sigma_\varepsilon^2$.

$$
\begin{aligned}
\ln L_{T-1}(\upsilon) &= -\frac{N(T-1)}{2}\ln(2\pi) - N(T-1)[\ln(\sigma_\mu^2) - \ln(\lambda)] - N\ln((T-1)\lambda+1) \\
&+ T\ln|I_N - \rho W| - \frac{1}{2(\sigma_\mu^2/\lambda)}e'\Omega^{-1}e \qquad (7) \\
\lambda &= \sigma_\mu^2/\sigma_\varepsilon^2 \\
e &= (Pg - Z\delta) \\
Z &= \begin{pmatrix} Z_1 & \dots & Z_{T-1} \end{pmatrix}' \\
Z_{t-1} &= \begin{pmatrix} \ln y_{t-1} & D\ln y_{t-1} & \iota_N & D\iota_N & X_{t-1} & (\tilde{D}\odot X_{t-1}) \end{pmatrix} \\
\delta &= \begin{pmatrix} \psi \\ \tilde{\psi} \\ \alpha \\ \tilde{\alpha} \\ \beta \\ \tilde{\beta} \end{pmatrix} \\
\upsilon &= (\phi, \rho, \theta, \lambda, \delta').
\end{aligned}
$$

For Bayesian model comparison purposes we wish to find an expression for the log-marginal likelihood. Zellner (1971) sets forth the basic Bayesian approach to model comparison. This involves specifying prior probabilities for each model as well as prior distributions for the regression parameters. Posterior model probabilities are calculated for each model and used for inferences regarding the "best model". The Bayesian theory behind model comparison involves specifying prior probabilities for each of the r alternative models $\{R_1, R_2, \dots, R_r\}$ under consideration, which we label $\pi(R_q), q = 1, \dots, r$, as well as prior distributions for the parameters $\pi(\upsilon)$. If the sample data are to determine the posterior model probabilities, the prior probabilities should be set to equal values of $1/r$, making each model equally likely a priori. We treat the spatial weight matrix W as fixed and exogenous, relying on a weight structure consisting of the 10 nearest neighboring regions (measured in terms of great circle distances). The motivation for this is that use of the 10 nearest neighboring regions allows the island regions of Greece to be connected to mainland Greece. We also treat the number of clubs as fixed at two, but future work will consider extending this.

The prior distributions for the parameters are combined with the likelihood for (g, Z, W) conditional on υ as well as the set of models R, which we denote $p(g|\upsilon, R, Z, W)$. The joint probability for R_q, υ, and g takes the form in (8), for the qth model based on a sample split at initial period income level $m = m_q$.

$$p(R_q, \upsilon, g, Z, W, m = m_q) = \pi(R_q)\pi(\upsilon|R_q)p(g|\upsilon, R, Z, W) \tag{8}$$

Application of Bayes rule produces the joint posterior for both models and parameters as:

$$p(R_q, \upsilon|g, Z, W) = \frac{\pi(R_q)\pi(\upsilon|R_q)p(g|\upsilon, R, Z, W)}{p(g)} \tag{9}$$

The posterior probabilities regarding the models take the form:

$$p(R_q|g, Z, W) = \int p(R_q, \upsilon|g, Z, W)d\upsilon \tag{10}$$

which requires integration over the parameter vector υ. We follow LeSage and Parent (2007) who develop expressions for the log-marginal likelihood in the case of a cross-sectional model by analytically integrating out the parameters δ and σ_ε, leaving a simple univariate numerical integration over the spatial dependence parameter ρ. Things are more complicated here, but we are able to analytically integrate out the parameters δ (see Appendix A for technical details). This requires that we fix $\lambda = \sigma_\mu^2/\sigma_\varepsilon^2$.

We make the following observation regarding λ. For small values of λ the effects magnitudes are likely to be close to their mean values of zero and not of substantive importance. Large values for the effects magnitudes accompanied by large values of λ likely suggest model specification problems. This leads us to posit that a well-specified model would exhibit model probabilities that should not be sensitive to fixing the value of λ, based on say, estimates for the parameters $\sigma_\mu^2, \sigma_\varepsilon^2$ from estimation of the panel data model with no dummy variables. We examine the resulting posterior model probabilities at values of $(1/2)\hat{\lambda}$ and

$2\hat{\lambda}$ as well as the estimated value: $\hat{\lambda} = \hat{\sigma}_\mu^2/\hat{\sigma}_\varepsilon^2$, to check robustness of results with regard to this ratio of variances.

Another simplification can be achieved by fixing the parameter $\theta = -\rho\phi$ which is a restriction implied by the space-time filter view of the panel data model specification. Parent and LeSage (2011) discuss the role of this restriction which simplifies both estimation and interpretation of the model. They also show that the restriction is often consistent with sample data sets, a finding for our empirical application as well. The advantage of this restriction is that we have a bivariate numerical integration problem involving the parameters ϕ and ρ rather than trivariate numerical integration.

Appendix B provides an illustration of this model comparison procedure based on a set of growth rates generated using sample data from our 216 European Union regions. Results presented in Appendix B show that the method performs well in identifying the model generated to have two regimes based on initial period income levels above and below 20,000. The appendix also explores sensitivity of the procedure to values of $(1/2)\hat{\lambda}$ and $2\hat{\lambda}$ rather than the estimated value. Appendix B relied on estimates from the growth relationship in (6), but altered values of ϕ and ρ so they did not obey the restriction $\theta = -\phi\rho$. This did not appear to produce erroneous inferences regarding the correct model.

3 Empirical club assignments

A description of the sample data used with the methodology described in Section 3.1. with the club assignment results reported in Section 3.2.

3.1 The sample data

Our sample is a cross-section of 216 regions representing the 15 pre-2004 EU member states, Norway and Switzerland over the 1995-2005 period. The units of observation are the NUTS-2 regions[10] (NUTS revision 2003). These regions, though varying in size, are generally considered to be appropriate spatial units for modelling and analysis purposes.

[10]We exclude the Spanish North African territories of Ceuta and Melilla, the Portuguese non-continental territories Azores and Madeira, the French Départements d'Outre-Mer Guadaloupe, Martinique, French Guayana and Réunion.

In most cases, they are sufficiently small to capture subnational variations. But we are aware that NUTS-2 regions are formal rather than functional regions, and their delineation does not represent the boundaries of regional growth processes very well. The sample regions include regions located in Western Europe covering Austria (nine regions), Belgium (11 regions), Denmark (one region), Finland (five regions), France (22 regions), Germany (40 regions), Greece (13 regions), Ireland (two regions) Italy (20 regions), Luxembourg (one region), the Netherlands (12 regions), Norway (seven regions), Portugal (five regions), Spain (16 regions), Sweden (eight regions), Switzerland (seven regions) and United Kingdom (37 regions).

We use gross-value added, GVA, rather than gross regional product at market prices as a proxy for regional income. The proxy is measured in accordance with the European Systems of Accounts (ESA) 1995. The data for the EU-regions come from Eurostat's Regio database, and those for Norway and Switzerland from Statistics Norway (Division for national Accounts) and the Swiss Office Féderal de la Statistique (Comptes Nationaux), respectively.

We use three variables in the dynamic space-time growth regression model to group regions based on initial levels: physical capital, knowledge capital and human capital. Physical capital stock data is not available in Cambridge econometrics database, but gross fixed capital formation in current prices is. Thus, the stocks of physical capital were derived for each region from investment flows, using the perpetual inventory method. We applied a constant rate of 10 percent depreciation, and the annual flows of fixed investments were deflated by national gross-fixed capital formation deflators. The mean annual rate of growth, which precedes the benchmark year 1995, covers the period 1990-1994 to estimate initial regional physical capital stocks.

Corporate patent applications are used to proxy knowledge capital. Corporate patents cover inventions of new and useful processes, machines, manufactures, and compositions of matter. To the extent that patents document inventions, an aggregation of patents is arguably more closely related to a stock of knowledge than is an aggregation of R&D expenditures. However, a well known problem of using patent data is that technological inventions are not all patented. This could be because of applying for a patent, is a strategic decision and, thus, not all patentable inventions are actually patented. Even if this is not an issue, as long as a large part of knowledge is tacit, patent statistics will necessarily miss that part, because codification is necessary for patenting to occur.

Patent stocks were derived from European Patent Office (EPO) documents. Each EPO document provides information on the inventor(s), his or her name and address, the company or institution to which property rights have been assigned, citations to previous patents, and a description of the device or process. To create the patent stocks for 1995-2005, the EPO patents with an application date 1990-2005 were transformed from individual patents into stocks by first sorting based on the year that a patent was applied for, and second the region where the inventor resides. In the case of cross-region inventor teams we used the procedure of fractional rather than full counting. Then for each region i, patent stocks were derived from patent data, using the perpetual inventory method. Because of evident complications in tracking obsolescence over time, we used a constant depreciation rate of 12 years that corresponds to the rate of knowledge obsolescence in the US over the past century, as found in Caballero and Jaffe (1993). Patent stocks were initialized the same way as physical capital.

There is no clear-cut consensus of how human capital should be represented and measured. In this study we follow Fischer et al. (2009) to measure human capital in terms of educational attainment based on data for the active population aged 15 years and older that attained the level of tertiary education, as defined by the International Standard Classification of Education (ISCED) 1997 classes five and six. This variable is clearly imperfect: it completely ignores primary and secondary education, and on-the-job training, and does not account for the quality of education.

3.2 Club assignments of the regions

Let us start by noting that most theoretical models of multiple steady states (see, for example, Azariadis and Drazen 1999; Galor 1996) predict that if (regional) economies are concentrated around several steady states, then their initial per capita output levels (here measured in terms of GVA per capita levels) will fall into distinct (i.e. non-overlapping) categories (Durlauf and Johnson 1995).

Figure 1 shows a frequency distribution of 1995 GVA per capita levels for regions where this was below 50,000.[11] In the figure, each bin of the histogram is 2,000 with the labels centered on these bins. There is a decline in the number of regions with 1995 GVA per capita levels beginning at 14,000. Another decline exists around 22,000 to 24,000, with an even more marked decline from 26,000 to 28,000. reflecting a smaller number of EU regions with initial period income levels above 16,000. Another decline exists around 22,000 to 24,000, with an even more marked decline from 26,000 to 28,000. Large changes in the number of regions that would arise from splitting the sample of regions at these income levels would lead to more dramatic changes in the posterior model probabilities. This should be clear by considering that adding or subtracting a single region from the set of Club 1 regions should lead to small changes in the log-marginal likelihood (and associated model probabilities). In contrast, changing the sub-samples through addition or subtraction of many regions would lead to larger changes in the posterior model probabilities.

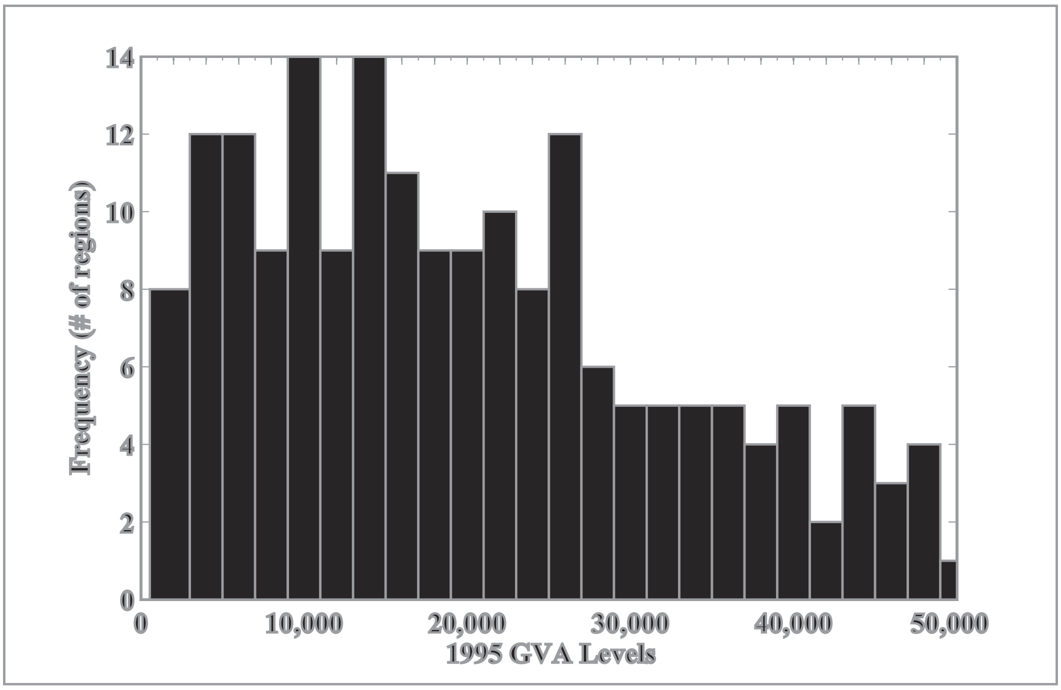

Figure 1: Frequency distribution of 1996 GVA per capita levels

We calculated posterior model probabilities for models based on a number of different

[11] This restriction was implemented to improve scaling of the figure.

candidate income levels $m_q, q = 1, \ldots, M$, based on a different level of initial period income. Each split of the regions into two clubs was based on differing initial period income levels implemented using dummy variable vectors, D_q, which can be treated as a separate model in the model comparison procedures. We use the analytical expressions derived in Section 4.1 for the log-marginal likelihood in conjunction with bivariate numerical integration over the parameters ϕ and ρ to find the log-marginal likelihoods that are required to calculate posterior model probabilities. The results are shown in Table 1 for a set of splits based on initial period income levels ranging from 8,000 to 30,000. Results are also shown for various values of the parameter λ, which was fixed at values ranging from 0.3 to 1, with a value of 0.4 indicated by estimates from both models (1) and (2) without dummy variables.

Table 1: Posterior model probabilities for various splits of the sample based on 1995 levels of income

Sample split y_0 levels	Model (1) Probs ($\lambda = 0.3$)	Model (1) Probs ($\lambda = 0.4$)	Model (1) Probs ($\lambda = 0.5$)	Model (1) Probs ($\lambda = 1$)
8,000	0.0206	0.0102	0.0739	0.0064
10,000	0.0251	0.0118	0.0214	0.0534
12,000	0.0063	0.0012	0.0011	0.0039
14,000	0.0260	0.0099	0.0089	0.0162
16,000	0.7165	0.5206	0.7062	0.4670
18,000	0.0013	0.0058	0.0109	0.0007
20,000	0.0070	0.0143	0.0039	0.0063
22,000	0.1777	0.4147	0.1634	0.1809
24,000	0.0150	0.0022	0.0065	0.0080
26,000	0.0009	0.0059	0.0004	0.0003
28,000	0.0031	0.0004	0.0008	0.2532
30,000	0.0006	0.0030	0.0026	0.0037
Sample split y_0 levels	Model (2) Probs ($\lambda = 0.3$)	Model (2) Probs ($\lambda = 0.4$)	Model (2) Probs ($\lambda = 0.5$)	Model (2) Probs ($\lambda = 1$)
8,000	0.0591	0.4197	0.1664	0.0442
10,000	0.0451	0.0535	0.0667	0.2515
12,000	0.0064	0.0089	0.0084	0.0046
14,000	0.4476	0.3962	0.3797	0.4338
16,000	0.3201	0.1038	0.1020	0.0719
18,000	0.0135	0.0036	0.0074	0.0030
20,000	0.0089	0.0022	0.0040	0.0063
22,000	0.0377	0.0071	0.0890	0.1828
24,000	0.0125	0.0017	0.0056	0.0002
26,000	0.0014	0.0010	0.1694	0.0008
28,000	0.0468	0.0004	0.0005	0.0003
30,000	0.0009	0.0021	0.0008	0.0006

There appears to be support for a split of the sample around 16,000, for the case of model

(1), with models based on this split level for initial period (1995) income levels exhibiting the highest posterior model probabilities. These results were relatively stable across values of the noise variance ratio parameter λ, always giving slightly more posterior probability support for a split at 16,000. It should be noted that we are forcing a choice of "the best model" from this finite set of models based on initial period income levels ranging from 8,000 to 30,000. This means that the posterior probabilities sum to unity, with all mass being assigned to the finite set of models. For the specification in model (2), results point to a split of the regions into two clubs based on 14,000 initial period income levels, which is close to the model (1) results. It should be noted that these results involve a smaller panel of only 8 periods because of the imposition of a 3-year lag on endowments.

The conclusion we draw is that the preponderance of evidence points to the existence of two clubs based on splitting the sample at initial period (1995) per capita GVA levels of 16,000. Figure 2 shows a map of the EU regions classified into the two clubs based on a split according to the 1995 GVA per capita levels for regions above and below 16,000.

4 Space-time dynamics for the two clubs

The second step of our approach involves estimating a space-time dynamic panel data model that uses (logged) levels of regional income as the dependent variable and (logged) levels of previous period endowments of physical, knowledge and human capital stocks, to examine the response of regional income levels over space and time to changes in initial period endowments, in each of the two clubs of regions. Our focus is on the partial derivative effects associated with changing the physical, knowledge and human capital stocks. Section 4.1 outlines the fixed effects variant of our dynamic space-time panel data model used for calculating dynamic response elasticities for regional income levels over space and time, to changes in initial period endowments of physical, knowledge and human capital stocks. Section 4.2 reports parameter estimates for the model along with scalar summary measures of the dynamic elasticity responses of income levels to changes in initial endowments.

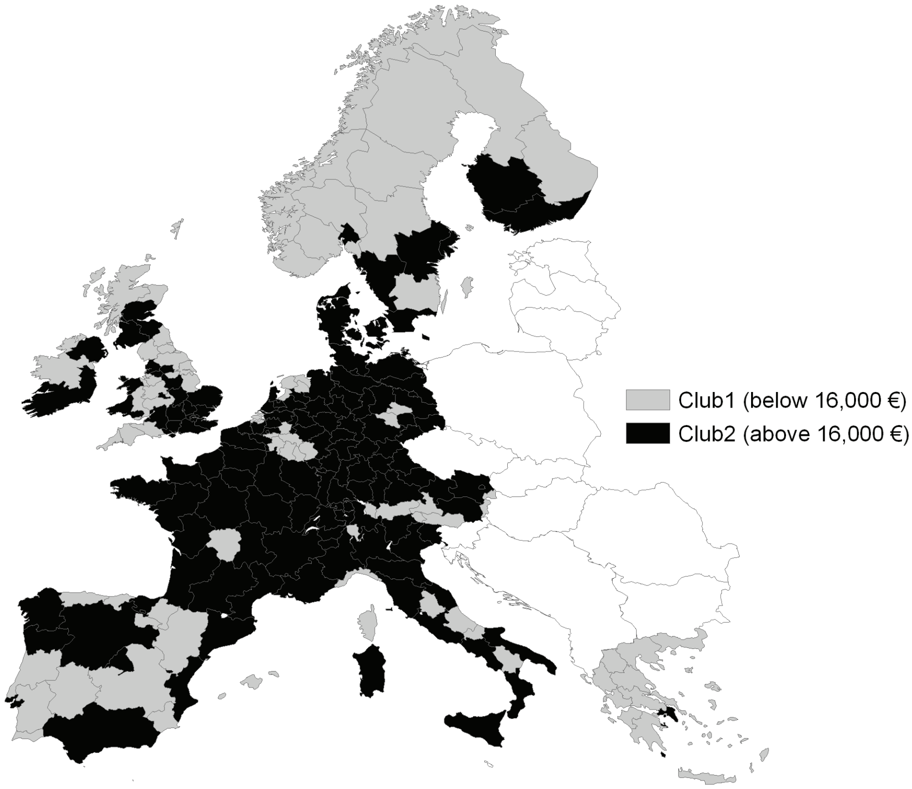

Figure 2: A map of regions classified into two clubs

4.1 The space-time levels relationship

We use a fixed effects variant of our dynamic space-time panel data model, and focus on the (logged) levels relationship between the dependent y_t and explanatory variables X_{t-1} as well as a linear combination of neighbouring region explanatory variables WX_{t-1}, so we can calculate dynamic response elasticities for regional income levels over space and time, to changes in initial period endowments of physical, knowledge and human capital stocks. (Ertur and Koch (2007) derive this type of expression where neighbouring region explanatory variables arise in a growth regression framework from neoclassical growth theory.) The (fixed effects) dynamic space-time panel model takes the form:

$$y_t \quad = \quad \phi y_{t-1} + \rho W y_t + \theta W y_{t-1} + X_{t-1}\beta + (\tilde{D} \odot X_{t-1})\tilde{\beta}$$

$$+ \quad WX_{t-1}\eta + (\tilde{D} \odot WX_{t-1})\tilde{\eta} + F\gamma + \varepsilon_t, \qquad t = 2, \ldots, T \tag{11}$$

where y_t, X_{t-1} have been log-transformed, ε_t is *i.i.d.* across i and t with zero mean and variance $\sigma_\varepsilon^2 I_N$, and F represent fixed effects with γ the associated parameters.

To explore the impact of using a 3-year time interval relationship on our estimates and inferences, we also estimate the model in (12).

$$\begin{aligned} y_t &= \phi y_{t-3} + \rho W y_t + \theta W y_{t-3} + X_{t-3}\beta + (\tilde{D} \odot X_{t-3})\tilde{\beta} \\ &+ \quad WX_{t-3}\eta + (\tilde{D} \odot WX_{t-3})\tilde{\eta} + F\gamma + \varepsilon_t, \qquad t = 4, \ldots, T \end{aligned} \tag{12}$$

We rely on a Bayesian Markov Chain Monte Carlo estimation scheme described in Debarsy, Ertur and LeSage (2012) to produce estimates of the parameters in the model. Our focus here is on the partial derivative effects associated with changing the explanatory variables in this model, reflecting human and physical capital stocks as well as knowledge stocks.

This model has own- and cross-partial derivatives that measure the impact on own- and other-regions income. We will use y_{it} to reference elements in the $N \times 1$ vector y_t pertaining to the ith element/region at time t, and we drop the explicit ln symbols for notational simplicity. The own-partial derivative: $\partial y_{it}/\partial X_{it}^k$, represents the time t *direct effect* on region i's (logged) income level (at time t), arising from a change in the kth explanatory variable (say logged physical capital levels) in region i (at time t). There is also a cross-partial derivative $\partial y_{jt}/\partial X_{it}^k$ that measures the time t *indirect effect*, that falling on regions (j) other than i, where most of the spatial spillover impacts fall on regions j that are nearby or neighbours to region i.

We are most interested in partial derivatives that measure how region i's (logged) income level responds over time to changes in the initial period (logged) endowment levels of physical and human capital, as well as knowledge stocks, since this is the essence of the debate concerning regional convergence in levels of income over time. The model allows us to calculate partial derivatives that can quantify the magnitude and timing of regional income responses at various time horizons to changes in the initial period levels of the explanatory variables. Expressions for these are presented and discussed in the sequel. We simply note here that we are referring to: $\partial y_{it+T}/\partial X_{it}^k$ which measures the $T-$horizon own-region i response to changes in its initial endowments, and $\partial y_{jt+T}/\partial X_{it}^k$, that reflects spillovers/diffusion effects over time that impact other regions j when region i's initial period human and physical capital or knowledge stocks are changed.

We follow Yu, de Jong and Lee (2008) and treat the dynamic space-time process as conditional on the initial cross-section. A careful analysis of issues related to treatment of the first period observation can be found in Parent and LeSage (2011), and we do not address this issue here. For simplicity of exposition, we assume that the first period is only subject to spatial dependence, which allows us to write the model as in (13), with accompanying definitions in (14), (15), (16) and (17).

$$
\begin{aligned}
QY &= \begin{pmatrix} X & WX \end{pmatrix} \begin{pmatrix} \beta \\ \eta \end{pmatrix} + (I_{T-1} \otimes \tilde{D}) \odot \begin{pmatrix} X & WX \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ \tilde{\eta} \end{pmatrix} + F\gamma + \varepsilon \qquad (13) \\
Q &= \begin{pmatrix} B & 0_{N\times N} & 0_{N\times N} & \dots & 0_{N\times N} \\ A & B & 0_{N\times N} & \dots & 0_{N\times N} \\ 0_{N\times N} & A & B & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0_{N\times N} \\ 0_{N\times N} & \dots & 0_{N\times N} & A & B \end{pmatrix}, \qquad (14) \\
A &= -(\phi I_N + \theta W) \qquad (15) \\
B &= (I_N - \rho W) \qquad (16) \\
F &= \iota_{T-1} \otimes I_N \qquad (17)
\end{aligned}
$$

The dependent variable vector $Y = (y_2', \dots, y_T')'$, consisting of $N \times 1$ vectors of cross-sectional observations for each time period $y_t = (y_{1t}, \dots, y_{Nt})'$. The matrix $X = (x_2', \dots, x_T')'$, so that x_t denotes the $N \times K$ matrix of (lagged) non-stochastic regressors at time t. We use X_{it}^k to reference elements associated with the kth variable for region i at time t. The matrix product $[(I_{t-1} \otimes \tilde{D}) \odot X]$ applies the club dummy variables to the explanatory variables X and WX for each time period, allowing for parameters β, η associated with Club 1, the low initial period income club and parameters $\beta + \tilde{\beta}, \eta + \tilde{\eta}$ for Club 2, the high initial period income club.

The matrix L represents the time lag operator $Ly_t = y_{t-1}$. The $N \times 1$ column vector γ represents fixed effects parameters, and the $N(T-1) \times N$ matrix F the associated regional indicator variables. The disturbance vector $\varepsilon = (\varepsilon_2', \ldots, \varepsilon_T')'$ with $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ assumed to be *i.i.d.* across i and t, with zero mean and variance σ^2. Spatial dependence is measured by the parameter ρ and time dependence is reflected in the scalar parameter ϕ, while the covariance between space and time is captured by the term $L \otimes W$ and associated parameter θ. The space filter matrix $B = (I_N - \rho W)$ is nonsingular, where the scalar spatial dependence parameter is ρ and the $N \times N$ matrix W is assumed to be a known *row stochastic* spatial weight matrix (exogenous with row-sums of unity and with zeros on the diagonal). This matrix defines the dependence between cross-sectional spatial units. We will also assume that W was created by row-normalizing our 10 nearest neighbors matrix, so that all eigenvalues are less than or equal to one. To address time-specific effects, we apply the time mean differencing matrix transformation $J = I_{T-1} \otimes (I_N - (1/N)\iota_N \iota_N'))$ to put each time period in deviations from the time mean form.[12]

The associated data generating process (DGP) shown in (18).

$$Y = Q^{-1} \left[\begin{pmatrix} X & WX \end{pmatrix} \begin{pmatrix} \beta \\ \eta \end{pmatrix} + (I_T \otimes \tilde{D}) \odot \begin{pmatrix} X & WX \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ \tilde{\eta} \end{pmatrix} + F\gamma + \varepsilon \right] \quad (18)$$

Of course, the values taken by the kth explanatory variable change with time periods so we need to further elaborate expression (18). For future reference we note that Debarsy, Ertur and LeSage (2012) show that the matrix Q^{-1} takes the form of a lower-triangular block matrix, containing blocks with $N \times N$ matrices. [13]

[12] This transformation is applied to Y and X as well as F and it obliterates the intercept term from the model. For clarity we do not include this in the notation regarding our discussion of the partial derivative impacts on y_{t+T} arising from changes in X_{it}, since it does not influence these.

[13] See Parent and LeSage (2010) for the special case that arises when the restriction $\theta = -\rho\phi$ is imposed.

$$Q^{-1} = \begin{pmatrix} B^{-1} & 0_{N\times N} & 0_{N\times N} & 0_{N\times N} & \dots & 0_{N\times N} \\ C_1 & B^{-1} & 0_{N\times N} & 0_{N\times N} & \dots & 0_{N\times N} \\ C_2 & C_1 & B^{-1} & 0_{N\times N} & \dots & 0_{N\times N} \\ C_3 & C_2 & C_1 & B^{-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0_{N\times N} \\ C_{T-1} & C_{T-2} & \dots & C_2 & C_1 & B^{-1} \end{pmatrix} \tag{19}$$

$$C_s = (-1)^s(B^{-1}A)^s B^{-1}, \quad s = 1, \dots, T-1$$

One implication of this is that we need only calculate the two $N \times N$ matrices: A and B^{-1} to analyze the partial derivative impacts for any time horizon T. This means we can use a panel involving say ten years to analyze the cumulative impacts arising from a permanent change in endowments at any time t extending to future horizons $t + T$. Of course, long horizons where T represents 50 or 100 years are of interest for regional growth and convergence issues.

The one-period-ahead impact of a (permanent) change the kth variable at time t for regions in Club 1 are shown in (20) and those for regions in Club 2 are in (21).

$$\begin{aligned} \partial Y_{t+1}/\partial X_t^k &= C_1[I_N\beta_k + W\eta_k] && (20) \\ &= -B^{-1}(\phi I_N + \theta W)B^{-1}[I_N\beta_k + W\eta_k] \\ \partial Y_{t+1}/\partial X_t^k &= C_1[I_N(\beta_k + \tilde{\beta}_k) + W(\eta_k + \tilde{\eta}_k)] && (21) \\ &= -B^{-1}(\phi I_N + \theta W)B^{-1}[I_N(\beta_k + \tilde{\beta}_k) + W(\eta_k + \tilde{\eta}_k)] \end{aligned}$$

More generally, the T-period-ahead (cumulative) impact arising from a permanent change at time t in X_t^k takes the form in (22) for regions in Club 1 and (23) for Club 2 regions. Note that we are cumulating down the columns (or rows) of the matrix in (19). For interpretative purposes we follow LeSage and Pace (2009) who note that the columns represent partial derivative changes arising from a change in a single region, whereas the rows reflect changes in all regions.

$$
\begin{aligned}
\partial Y_{t+T}/\partial X_t^k &= \sum_{s=1}^{T} C_s[I_N\beta_k + W\eta_k] && (22)\\
\partial Y_{t+T}/\partial X_t^k &= \sum_{s=1}^{T} C_s[I_N(\beta_k + \tilde{\beta}_k) + W(\eta_k + \tilde{\eta}_k)] && (23)\\
C_s &= (-1)^s(B^{-1}A)^s B^{-1}
\end{aligned}
$$

By analogy to LeSage and Pace (2009), the main diagonal elements of the $N \times N$ matrix sums for time horizon T represent (cumulative) own-region impacts that arise from both time and spatial dependence. The off-diagonal elements of these matrix sums reflect diffusion over space and time. We note that it is not possible to separate out the time from space and space-time diffusion effects in this model.[14]

4.2 Dynamic elasticity responses for the two clubs

We first report parameter estimates for the model, although these are not directly interpretable in terms of the space-time dynamic impacts associated with changes in the explanatory variables on the dependent variable (regional income levels). Posterior means, medians and standard deviations as well as a ratio of the mean/standard deviation are reported for the space-time dependence parameters ϕ, ρ, θ and the noise variance parameter σ_ε^2 in Table 2.

From the table we see significant time, space and space-time dependence, with the restriction that $\theta = -\rho\phi$ discussed in Parent and LeSage (2011) being quite consistent with this dataset, since $0.7253 \times -0.78597 = -0.5701$, which is very close to the unrestricted estimate for $\theta = -0.5745$. In fact, the difference of 0.0044 between these two estimates is much smaller than the estimated standard deviation for θ equal to 0.0148. We come to a similar conclusion regarding the restriction that $\theta = -\rho\phi$ using the posterior medians in place of the means.

The table also reports coefficient estimates for the three explanatory variables used:

[14]See Parent and LeSage (2010) for the special case where space and time are separable.

Table 2: Dynamic space-time panel data model estimates

Posterior statistics	ϕ	ρ	θ	σ_ε^2
mean	0.72537	0.78598	-0.57451	0.000658
median	0.72701	0.78556	-0.57456	0.000658
std	0.00897	0.01214	0.01484	0.000021
mean/std	80.85365	64.72646	-38.69733	31.185820
Variables	Mean	Std	Mean/Std	$t-$probability
β capital stock	0.05242	0.0140	3.72	0.00020
β knowledge stock	0.00915	0.0061	1.49	0.13582
β human capital	0.03497	0.0093	3.73	0.00019
η W (capital stock)	0.00316	0.0242	0.13	0.89603
η W (knowledge stock)	0.02153	0.0101	2.12	0.03356
η W (human capital)	-0.07467	0.0202	-3.67	0.00024
$\tilde{\beta}$ capital stock	0.01265	0.0199	0.63	0.52684
$\tilde{\beta}$ knowledge stock	0.01670	0.0084	1.96	0.04918
$\tilde{\beta}$ human capital	-0.04969	0.0139	-3.55	0.00038
$\tilde{\eta}$ W (capital stock)	-0.04115	0.0295	-1.39	0.16338
$\tilde{\eta}$ W (knowledge stock)	-0.04624	0.0112	-4.10	0.00004
$\tilde{\eta}$ W (human capital)	0.10511	0.0263	3.98	0.00007

(logged) regional levels of physical capital stocks, knowledge stocks and human capital as β coefficients, along with those from an average of the 10 neighboring regions recorded as η coefficients on the WX variables. The coefficients for the Club 2 dummy variables associated with these two sets of explanatory variables are denoted using $\tilde{\beta}, \tilde{\eta}$, and we note that neighboring region η coefficients for knowledge stocks and human capital appear to exert a significant influence, as do the neighboring region dummy variable coefficients $\tilde{\eta}$ for knowledge stocks and human capital. The role of neighboring region endowments was ignored by Yu and Lee (2012) in their implementation of this model for US data. It should be noted that none of these coefficients $(\beta, \eta, \tilde{\beta}, \tilde{\eta})$ are directly interpretable as indicating how the dependent variable responds to changes in the explanatory variables, a point that has frequently been overlooked in the dynamic panel data model literature.

The dynamic elasticity responses are shown in Table 3 for the direct (own-region) responses to changes in the physical capital stock variable for both clubs. The direct effects estimates reported show time horizon zero effects that reflect simultaneous own-region spatial effects, while time horizons one to 20 years include the future period own-region impacts that arise from time dependence as well as some spatiotemporal feedback effects. Note that in this model regional income depends on neighboring regions implying that future period changes in neighboring regions' income will set in motion a feedback loop that produces second order benefits to the own-region as a result of spatial spillover benefits generated for neighbors in earlier time periods.

The first column shows the time horizon $(t+T)$, while the second and third columns present the point estimates for the cumulative and marginal direct effects. The second column shows cumulative effects whereas the third column shows the marginal effect or period-by-period change. A mean divided by the empirically calculated standard deviation was used to produce a $t-$statistic and associated $p-$level reported in the fourth and fifth columns of the table, as a test of significance for the marginal effects estimates. This allows us to see when the period-by-period response dies down to become insignificantly different from zero. It should be clear that the (marginal) response (over space and time) to a sustained or permanent shock in the physical capital variable dies down to zero, consistent with the fact that our model estimates lie in the region of stability ($\phi+\rho+\theta<1$, see Parent and LeSage 2011 for a derivation and discussion of these conditions). The same format was used to report direct effects for Club 2 alongside those of Club 1 for comparison purposes.

The dynamic elasticity responses reveal that a 10 percent increase in physical capital stocks in Club 1 (low-income) regions would lead to a long-run increase in income (GVA per capita) of 2.2 percent, and a very similar 2.4 percent increase for Club 2 (high-income) regions. The mean/standard deviations calculated for the marginal responses shows that increases in physical capital have a long-lived impact on regional incomes, since the marginal effects are significantly different from zero (using the 95% level of significance) out to a 21- and 27-year time horizon for Club 1 and Club 2, respectively. These results suggest no difference in how low and high income regions are able to convert increased physical capital stocks into higher regional income levels.

Table 4 shows the indirect (spatial spillover) effects associated with a change in physical capital stocks, using the same format as in Table 3. Here we see significant positive spillovers for Club 1 regions that extend out to around a 4-year time horizon (using the 95% level). The cumulative spillover magnitude of 0.78 appear very large when compared to the direct effects magnitude of 0.22, but these are cumulative spillovers, where the cumulation takes place over all neighboring regions, neighbors to the neighboring regions and so on. Effects

Table 3: Dynamic elasticity direct responses for changes in physical capital

	Club 1				Club 2			
Periods	Cumulative	Marginal	$t-$stat	$t-$prob	Cumulative	Marginal	$t-$stat	$t-$prob
0	0.0632	0.0632	4.77	0.0000	0.0691	0.0691	5.14	0.0000
1	0.1084	0.0452	4.77	0.0000	0.1186	0.0494	5.08	0.0000
2	0.1408	0.0324	4.74	0.0000	0.1540	0.0354	5.00	0.0000
3	0.1641	0.0232	4.71	0.0000	0.1795	0.0254	4.92	0.0000
4	0.1808	0.0166	4.65	0.0000	0.1977	0.0182	4.82	0.0000
5	0.1927	0.0119	4.59	0.0000	0.2108	0.0130	4.73	0.0000
6	0.2013	0.0086	4.51	0.0000	0.2202	0.0093	4.62	0.0000
7	0.2075	0.0061	4.42	0.0000	0.2269	0.0067	4.52	0.0000
8	0.2120	0.0044	4.32	0.0000	0.2318	0.0048	4.41	0.0000
9	0.2152	0.0032	4.20	0.0000	0.2352	0.0034	4.30	0.0000
10	0.2175	0.0023	4.07	0.0000	0.2377	0.0025	4.19	0.0000
11	0.2191	0.0016	3.93	0.0000	0.2396	0.0018	4.08	0.0000
12	0.2204	0.0012	3.78	0.0001	0.2409	0.0012	3.97	0.0000
13	0.2212	0.0008	3.62	0.0003	0.2418	0.0009	3.86	0.0001
14	0.2219	0.0006	3.44	0.0005	0.2425	0.0006	3.75	0.0001
15	0.2223	0.0004	3.27	0.0010	0.2429	0.0004	3.64	0.0002
16	0.2227	0.0003	3.08	0.0020	0.2433	0.0003	3.52	0.0004
17	0.2229	0.0002	2.89	0.0037	0.2436	0.0002	3.41	0.0006
18	0.2231	0.0001	2.70	0.0068	0.2437	0.0001	3.29	0.0010
19	0.2232	0.0001	2.51	0.0118	0.2439	0.0001	3.17	0.0015
20	0.2233	0.0000	2.33	0.0197	0.2440	0.0000	3.05	0.0023
21	0.2234	0.0000	2.15	0.0314	0.2440	0.0000	2.92	0.0034
22	0.2234	0.0000	1.97	0.0479	0.2441	0.0000	2.79	0.0052
23	0.2235	0.0000	1.81	0.0698	0.2441	0.0000	2.65	0.0079
24	0.2235	0.0000	1.65	0.0973	0.2442	0.0000	2.51	0.0119
25	0.2235	0.0000	1.51	0.1302	0.2442	0.0000	2.36	0.0179
26	0.2235	0.0000	1.37	0.1679	0.2442	0.0000	2.22	0.0264
27	0.2235	0.0000	1.25	0.2091	0.2442	0.0000	2.07	0.0384
28	0.2235	0.0000	1.14	0.2527	0.2442	0.0000	1.92	0.0547
29	0.2235	0.0000	1.04	0.2975	0.2442	0.0000	1.77	0.0762
30	0.2235	0.0000	0.94	0.3422	0.2442	0.0000	1.62	0.1034

falling on any individual region are smaller than the direct effects, consistent with spillovers being a "second order effect". This can be seen by considering that there are 10 first order neighbors alone, so if we divide the spillover/indirect effects estimates by a factor of 10, the marginal impacts of 0.078 associated with a single region are much smaller than the direct effects. Further note that we should in reality divide by a number much greater than the 10 first order neighbors, since these effects emanate out to more distant neighbors as time passes, a phenomenon representing spatial diffusion impacts. See Parent and LeSage (2009) for a decomposition of the effects into time-specific and space-specific as well as diffusion-specific impacts.

In contrast to the Club 1 significant spatial spillovers to neighboring regions, Club 2 regions exhibit no significant indirect (spatial spillover) effects. The implication is that increases in physical capital stock for Club 2 regions leads to high income levels for the region itself, but does not influence current or future income levels of neighboring regions. There may of course be differences in the type of capital stock changes taking place in higher income (Club 2) regions than in lower income (Club 1) regions. For example, increases in physical capital representing shared resources such as public transportation infrastructure would be more likely to produce spatial spillovers than physical capital put in place by private firms.

Table 4: Dynamic elasticity indirect responses for changes in physical capital

	Club 1				Club 2			
Periods	Cumulative	Marginal	$t-$stat	$t-$prob	Cumulative	Marginal	$t-$stat	$t-$prob
0	0.2190	0.2190	2.70	0.0069	0.0824	0.0824	0.89	0.3731
1	0.3733	0.1542	2.85	0.0044	0.1385	0.0561	0.91	0.3603
2	0.4829	0.1095	2.78	0.0053	0.1771	0.0386	0.92	0.3557
3	0.5613	0.0784	2.53	0.0112	0.2040	0.0268	0.91	0.3604
4	0.6179	0.0566	2.21	0.0267	0.2230	0.0189	0.88	0.3746
5	0.6592	0.0412	1.91	0.0560	0.2365	0.0135	0.84	0.3963
6	0.6895	0.0302	1.65	0.0985	0.2463	0.0097	0.80	0.4231
7	0.7119	0.0224	1.44	0.1495	0.2535	0.0071	0.75	0.4520
8	0.7286	0.0167	1.27	0.2035	0.2588	0.0053	0.70	0.4808
9	0.7411	0.0125	1.13	0.2565	0.2628	0.0039	0.66	0.5081
10	0.7507	0.0095	1.02	0.3064	0.2658	0.0030	0.62	0.5331
11	0.7579	0.0072	0.93	0.3521	0.2681	0.0023	0.58	0.5556
12	0.7635	0.0055	0.85	0.3933	0.2699	0.0017	0.55	0.5757
13	0.7677	0.0042	0.78	0.4302	0.2713	0.0013	0.53	0.5935
14	0.7711	0.0033	0.73	0.4632	0.2724	0.0010	0.51	0.6096
15	0.7737	0.0025	0.68	0.4928	0.2732	0.0008	0.49	0.6240
16	0.7757	0.0020	0.64	0.5193	0.2739	0.0006	0.47	0.6373
17	0.7773	0.0015	0.60	0.5431	0.2745	0.0005	0.45	0.6494
18	0.7785	0.0012	0.57	0.5647	0.2749	0.0004	0.43	0.6608
19	0.7795	0.0010	0.54	0.5843	0.2752	0.0003	0.42	0.6714
20	0.7803	0.0007	0.52	0.6022	0.2755	0.0002	0.41	0.6815
21	0.7810	0.0006	0.49	0.6185	0.2757	0.0002	0.39	0.6911
22	0.7815	0.0005	0.47	0.6336	0.2759	0.0001	0.38	0.7004
23	0.7819	0.0004	0.45	0.6476	0.2761	0.0001	0.37	0.7092
24	0.7822	0.0003	0.43	0.6606	0.2762	0.0001	0.36	0.7178
25	0.7825	0.0002	0.42	0.6726	0.2763	0.0000	0.35	0.7261
26	0.7827	0.0002	0.40	0.6840	0.2764	0.0000	0.33	0.7341
27	0.7829	0.0001	0.39	0.6946	0.2764	0.0000	0.32	0.7418
28	0.7830	0.0001	0.37	0.7046	0.2765	0.0000	0.31	0.7493
29	0.7831	0.0001	0.36	0.7140	0.2765	0.0000	0.31	0.7565
30	0.7832	0.0000	0.35	0.7230	0.2765	0.0000	0.30	0.7635

An implication of these results is that Club 1 (low income) regions that are close neighbours to Club 2 (high income) regions would not benefit greatly from spatial spillovers and diffusion effects arising from increases of physical capital stocks in Club 2 regions. In contrast, Club 2 (and Club 1) regions would benefit from spillover and diffusion effects as a result of being neighbours to Club 1 regions where physical capital stocks are increasing.

Analysis of the total (cumulative) dynamic response elasticities shows that changes in physical capital stocks for the Club 1 regions (those with low initial period incomes) produce a long-run response of 10 percent higher level of income in response to a 10 percent increase in initial period physical capital stock. This represents a 2.2 percent direct or own-region impact and a 7.8 percent cumulative spatial and space-time diffusion impact. In contrast, regions in Club 2 (those with higher initial period incomes) exhibited a total (cumulative) long-run response of 2.4 percent to a 10 percent increase in initial period physical capital stock, with the difference between Club 1 and 2 regions accounted for by the lack of spatial spillover impacts.

Table 5 shows the direct effect responses to changes in knowledge stocks for regions in Clubs 1 and 2. Here we see (cumulative long-run) direct response for the Club 2 regions (0.0948) that is almost double that for Club 1 regions (0.0563). The impact of increased knowledge stocks also lasts about twice as long for Club 2 regions (25 years) versus Club 1 (11 years), based on the calculated $t-$statistics and 95% level of significance for the marginal effects. This would indicate that Club 2 regions benefit more from increased knowledge stocks than Club 1 regions, and that knowledge stocks survive longer in Club 2 regions. This is perhaps the result of using outdated knowledge/technology in the lower income regions. The magnitude of direct dynamic elasticity response of regional income levels to increased knowledge stocks is less than the response to increased physical capital for both clubs: for Club 1, 0.0563 (knowledge capital) versus 0.22 (physical capital) and for Club 2, 0.0948 (knowledge capital) versus 0.24 (physical capital).

Indirect effects responses are shown in Table 6, where we see the same pattern as in the case of physical capital, where indirect (spatial spillover) effects are not significantly different from zero for the Club 2 regions. There are significant marginal effects estimates for Club 1 regions that last four years (using the 95% level). The long-run magnitude of

Table 5: Dynamic elasticity direct responses for changes in knowledge stocks

	Club 1				Club 2			
Periods	Cumulative	Marginal	$t-$stat	$t-$prob	Cumulative	Marginal	$t-$stat	$t-$prob
0	0.0158	0.0158	2.63	0.0083	0.0267	0.0267	4.97	0.0000
1	0.0271	0.0113	2.69	0.0071	0.0459	0.0191	5.09	0.0000
2	0.0352	0.0081	2.73	0.0063	0.0596	0.0137	5.19	0.0000
3	0.0410	0.0058	2.75	0.0059	0.0695	0.0098	5.29	0.0000
4	0.0452	0.0041	2.75	0.0059	0.0766	0.0070	5.37	0.0000
5	0.0482	0.0030	2.72	0.0064	0.0817	0.0050	5.44	0.0000
6	0.0504	0.0021	2.67	0.0074	0.0854	0.0036	5.49	0.0000
7	0.0520	0.0015	2.60	0.0093	0.0880	0.0026	5.51	0.0000
8	0.0532	0.0011	2.50	0.0124	0.0899	0.0018	5.51	0.0000
9	0.0540	0.0008	2.38	0.0173	0.0913	0.0013	5.47	0.0000
10	0.0546	0.0006	2.24	0.0247	0.0922	0.0009	5.41	0.0000
11	0.0551	0.0004	2.10	0.0354	0.0929	0.0007	5.32	0.0000
12	0.0554	0.0003	1.95	0.0503	0.0934	0.0005	5.19	0.0000
13	0.0556	0.0002	1.81	0.0699	0.0938	0.0003	5.04	0.0000
14	0.0558	0.0001	1.67	0.0944	0.0941	0.0002	4.85	0.0000
15	0.0559	0.0001	1.54	0.1236	0.0943	0.0001	4.64	0.0000
16	0.0560	0.0000	1.41	0.1568	0.0944	0.0001	4.41	0.0000
17	0.0561	0.0000	1.30	0.1932	0.0945	0.0001	4.16	0.0000
18	0.0562	0.0000	1.19	0.2316	0.0946	0.0000	3.89	0.0001
19	0.0562	0.0000	1.10	0.2711	0.0946	0.0000	3.62	0.0002
20	0.0562	0.0000	1.01	0.3107	0.0947	0.0000	3.34	0.0008
21	0.0562	0.0000	0.93	0.3497	0.0947	0.0000	3.07	0.0021
22	0.0563	0.0000	0.86	0.3874	0.0947	0.0000	2.80	0.0050
23	0.0563	0.0000	0.80	0.4235	0.0947	0.0000	2.54	0.0108
24	0.0563	0.0000	0.74	0.4576	0.0947	0.0000	2.30	0.0212
25	0.0563	0.0000	0.69	0.4897	0.0947	0.0000	2.07	0.0378
26	0.0563	0.0000	0.64	0.5195	0.0947	0.0000	1.86	0.0620
27	0.0563	0.0000	0.60	0.5472	0.0948	0.0000	1.67	0.0942
28	0.0563	0.0000	0.56	0.5729	0.0948	0.0000	1.49	0.1339
29	0.0563	0.0000	0.52	0.5965	0.0948	0.0000	1.34	0.1799
30	0.0563	0.0000	0.49	0.6182	0.0948	0.0000	1.20	0.2302

spillovers (0.64) is similar to that for physical capital (0.78), and the same caveat applies to interpreting this cumulative spillover magnitude for a single neighbouring region.

The total (cumulative) impact arising from changes in knowledge stocks for Club 1 regions (direct plus indirect) was 0.70, consisting of 0.0563 direct and 0.6441 indirect effects. For Club 2 regions total (cumulative) impact was smaller at 0.0948, consisting entirely of the direct impact. These imply that a 10 percent increase in knowledge stocks would lead to 7 percent higher (long-run) incomes in Club 1 regions and a 0.94 percent increase in Club 2 regions. The (long-run) regional income responses to an increase in knowledge stocks of 10 percent is smaller than to a 10 percent increase in physical capital, 7 versus 10 percent

Table 6: Dynamic elasticity indirect responses for changes in knowledge stocks

	Club 1				Club 2			
Periods	Cumulative	Marginal	$t-$stat	$t-$prob	Cumulative	Marginal	$t-$stat	$t-$prob
0	0.1740	0.1740	6.39	0.0000	-0.0090	-0.0090	-0.31	0.7540
1	0.2984	0.1243	5.05	0.0000	-0.0152	-0.0061	-0.28	0.7734
2	0.3880	0.0895	3.72	0.0002	-0.0192	-0.0040	-0.25	0.8010
3	0.4529	0.0649	2.84	0.0045	-0.0217	-0.0025	-0.20	0.8345
4	0.5004	0.0474	2.26	0.0234	-0.0233	-0.0015	-0.16	0.8714
5	0.5354	0.0349	1.87	0.0606	-0.0241	-0.0008	-0.11	0.9094
6	0.5613	0.0259	1.59	0.1105	-0.0245	-0.0004	-0.06	0.9467
7	0.5807	0.0193	1.38	0.1654	-0.0247	-0.0001	-0.02	0.9819
8	0.5953	0.0145	1.22	0.2202	-0.0246	0.0000	0.01	0.9857
9	0.6063	0.0110	1.09	0.2721	-0.0244	0.0001	0.05	0.9567
10	0.6147	0.0084	0.99	0.3199	-0.0242	0.0002	0.08	0.9311
11	0.6212	0.0064	0.90	0.3631	-0.0239	0.0002	0.11	0.9089
12	0.6261	0.0049	0.83	0.4021	-0.0237	0.0002	0.13	0.8897
13	0.6300	0.0038	0.77	0.4372	-0.0234	0.0002	0.15	0.8735
14	0.6330	0.0029	0.72	0.4687	-0.0232	0.0002	0.17	0.8598
15	0.6353	0.0023	0.67	0.4971	-0.0230	0.0001	0.19	0.8485
16	0.6372	0.0018	0.63	0.5228	-0.0229	0.0001	0.20	0.8392
17	0.6386	0.0014	0.60	0.5461	-0.0227	0.0001	0.21	0.8317
18	0.6398	0.0011	0.57	0.5673	-0.0226	0.0001	0.22	0.8257
19	0.6407	0.0009	0.54	0.5867	-0.0225	0.0001	0.22	0.8211
20	0.6414	0.0007	0.51	0.6045	-0.0224	0.0000	0.23	0.8176
21	0.6420	0.0005	0.49	0.6210	-0.0223	0.0000	0.23	0.8151
22	0.6425	0.0004	0.47	0.6362	-0.0222	0.0000	0.23	0.8134
23	0.6428	0.0003	0.45	0.6503	-0.0222	0.0000	0.23	0.8125
24	0.6431	0.0003	0.43	0.6634	-0.0221	0.0000	0.23	0.8122
25	0.6434	0.0002	0.41	0.6756	-0.0221	0.0000	0.23	0.8123
26	0.6436	0.0001	0.40	0.6871	-0.0220	0.0000	0.23	0.8129
27	0.6437	0.0001	0.38	0.6979	-0.0220	0.0000	0.23	0.8139
28	0.6439	0.0001	0.37	0.7080	-0.0220	0.0000	0.23	0.8152
29	0.6440	0.0001	0.36	0.7176	-0.0220	0.0000	0.23	0.8167
30	0.6441	0.0000	0.34	0.7266	-0.0219	0.0000	0.22	0.8184

for Club 1, and 0.94 versus 2.4 percent for Club 2.

Finally, the direct and indirect dynamic elasticity responses of regional income to changes in human capital are shown in Table 7 and Table 8, respectively. Here we see that Club 1 regions benefit from increases in own-region human capital, with the long-run elasticity of response being 0.093, whereas increases in human capital for Club 2 regions does not have a significant impact on region income levels. The positive impact for Club 1 regions is significant over a 12 year horizon (based on 95% significance), and the long-run magnitude is about twice that of knowledge capital and half that of physical capital.

The Club 1 indirect (spatial spillover) impacts are negative and significant out to a 7

Table 7: Dynamic elasticity direct responses for changes in human capital stocks

	Club 1				Club 2			
Periods	Cumulative	Marginal	$t-$stat	$t-$prob	Cumulative	Marginal	$t-$stat	$t-$prob
0	0.0263	0.0263	3.18	0.0014	-0.0136	-0.0136	-1.48	0.1388
1	0.0452	0.0189	3.24	0.0011	-0.0233	-0.0096	-1.46	0.1420
2	0.0589	0.0136	3.27	0.0010	-0.0302	-0.0069	-1.45	0.1464
3	0.0687	0.0097	3.26	0.0011	-0.0351	-0.0049	-1.43	0.1522
4	0.0757	0.0070	3.22	0.0012	-0.0386	-0.0034	-1.40	0.1596
5	0.0807	0.0050	3.15	0.0016	-0.0411	-0.0024	-1.37	0.1686
6	0.0843	0.0036	3.06	0.0022	-0.0428	-0.0017	-1.34	0.1796
7	0.0869	0.0025	2.93	0.0033	-0.0441	-0.0012	-1.30	0.1928
8	0.0888	0.0018	2.79	0.0052	-0.0450	-0.0008	-1.25	0.2084
9	0.0901	0.0013	2.63	0.0084	-0.0456	-0.0006	-1.20	0.2269
10	0.0910	0.0009	2.46	0.0137	-0.0460	-0.0004	-1.15	0.2485
11	0.0917	0.0006	2.29	0.0220	-0.0463	-0.0003	-1.09	0.2735
12	0.0922	0.0004	2.11	0.0348	-0.0466	-0.0002	-1.03	0.3022
13	0.0925	0.0003	1.93	0.0533	-0.0467	-0.0001	-0.96	0.3349
14	0.0927	0.0002	1.75	0.0790	-0.0468	-0.0001	-0.89	0.3716
15	0.0929	0.0001	1.58	0.1129	-0.0469	-0.0000	-0.81	0.4123
16	0.0930	0.0001	1.42	0.1556	-0.0469	-0.0000	-0.74	0.4568
17	0.0931	0.0000	1.26	0.2069	-0.0470	-0.0000	-0.66	0.5047
18	0.0932	0.0000	1.11	0.2660	-0.0470	-0.0000	-0.58	0.5553
19	0.0932	0.0000	0.97	0.3313	-0.0470	-0.0000	-0.51	0.6078
20	0.0932	0.0000	0.83	0.4010	-0.0470	-0.0000	-0.43	0.6610
21	0.0933	0.0000	0.71	0.4729	-0.0470	-0.0000	-0.36	0.7140
22	0.0933	0.0000	0.60	0.5450	-0.0470	-0.0000	-0.29	0.7657
23	0.0933	0.0000	0.50	0.6154	-0.0470	-0.0000	-0.23	0.8151
24	0.0933	0.0000	0.40	0.6828	-0.0470	-0.0000	-0.17	0.8615
25	0.0933	0.0000	0.32	0.7460	-0.0470	-0.0000	-0.12	0.9042
26	0.0933	0.0000	0.24	0.8043	-0.0470	-0.0000	-0.07	0.9430
27	0.0933	0.0000	0.17	0.8573	-0.0470	-0.0000	-0.02	0.9777
28	0.0933	0.0000	0.11	0.9050	-0.0470	0.0000	0.01	0.9917
29	0.0933	0.0000	0.06	0.9473	-0.0470	0.0000	0.04	0.9651
30	0.0933	0.0000	0.01	0.9847	-0.0470	0.0000	0.07	0.9423

year horizon, suggesting that higher human capital stocks in a Club 1 region i lead to lower income levels in neighboring regions j. The long-run magnitude of (cumulative spillover) impact is -0.96, but the impact falling on any single neighboring region would be much smaller for the reasons indicated in our discussion of physical capital stock cumulative spatial spillovers. This might suggests that increases in human capital stock in Club 1 regions take place at the expense of neighboring regions, which would be consistent with educated workers moving to nearby regions for employment.

There were positive but not significant spatial spillover impacts associated with changes in human capital stocks in Club 2 regions.

Table 8: Dynamic elasticity indirect responses for changes in human capital stocks

	Club 1				Club 2			
Periods	Cumulative	Marginal	$t-$stat	$t-$prob	Cumulative	Marginal	$t-$stat	$t-$prob
0	-0.2692	-0.2692	-3.34	0.0008	0.0836	0.0836	1.26	0.2048
1	-0.4601	-0.1908	-3.48	0.0004	0.1454	0.0617	1.25	0.2103
2	-0.5961	-0.1360	-3.44	0.0005	0.1912	0.0458	1.22	0.2210
3	-0.6936	-0.0974	-3.23	0.0012	0.2253	0.0341	1.18	0.2364
4	-0.7638	-0.0702	-2.91	0.0036	0.2509	0.0255	1.13	0.2557
5	-0.8148	-0.0509	-2.56	0.0103	0.2701	0.0192	1.08	0.2782
6	-0.8519	-0.0371	-2.24	0.0249	0.2846	0.0145	1.03	0.3028
7	-0.8792	-0.0272	-1.96	0.0496	0.2956	0.0110	0.97	0.3290
8	-0.8993	-0.0201	-1.72	0.0842	0.3040	0.0083	0.92	0.3559
9	-0.9143	-0.0149	-1.52	0.1262	0.3104	0.0064	0.87	0.3830
10	-0.9254	-0.0111	-1.36	0.1726	0.3154	0.0049	0.82	0.4098
11	-0.9338	-0.0083	-1.22	0.2205	0.3192	0.0038	0.77	0.4361
12	-0.9401	-0.0063	-1.10	0.2677	0.3222	0.0029	0.73	0.4615
13	-0.9449	-0.0047	-1.00	0.3128	0.3245	0.0023	0.69	0.4860
14	-0.9486	-0.0036	-0.92	0.3551	0.3263	0.0018	0.65	0.5094
15	-0.9514	-0.0028	-0.85	0.3942	0.3277	0.0014	0.62	0.5316
16	-0.9535	-0.0021	-0.78	0.4300	0.3288	0.0011	0.59	0.5528
17	-0.9552	-0.0016	-0.73	0.4626	0.3297	0.0008	0.56	0.5728
18	-0.9565	-0.0013	-0.68	0.4923	0.3304	0.0006	0.53	0.5917
19	-0.9575	-0.0010	-0.64	0.5192	0.3309	0.0005	0.51	0.6096
20	-0.9583	-0.0008	-0.60	0.5436	0.3314	0.0004	0.48	0.6265
21	-0.9590	-0.0006	-0.57	0.5657	0.3317	0.0003	0.46	0.6423
22	-0.9595	-0.0005	-0.54	0.5859	0.3320	0.0002	0.44	0.6573
23	-0.9599	-0.0003	-0.51	0.6043	0.3322	0.0002	0.42	0.6714
24	-0.9602	-0.0003	-0.49	0.6210	0.3324	0.0001	0.40	0.6848
25	-0.9604	-0.0002	-0.47	0.6364	0.3326	0.0001	0.38	0.6973
26	-0.9606	-0.0002	-0.45	0.6505	0.3327	0.0001	0.37	0.7091
27	-0.9608	-0.0001	-0.43	0.6635	0.3328	0.0000	0.35	0.7203
28	-0.9609	-0.0001	-0.41	0.6756	0.3329	0.0000	0.34	0.7308
29	-0.9610	-0.0001	-0.40	0.6867	0.3329	0.0000	0.33	0.7407
30	-0.9611	-0.0000	-0.38	0.6971	0.3330	0.0000	0.31	0.7501

5 Concluding remarks

This paper describes a two-step approach to identifying and interpreting regional convergence clubs in Europe. The first step uses a formal Bayesian model comparison methodology to classify the European regions into convergence clubs. Each region must be classified into one of M clubs. The classification takes place conditional on a space-time panel data model of regional income growth. Since observations are regions in our model the comparison problem is one of comparing models based on different assignments of each observation (region) to one of the q club categories based on initial period income levels.

Even for the case of $q = 2$, the classification problem leads to a high dimensional model space consisting of 2^N possible models where N is the number of regions in the sample that need to be compared. We use a procedure that splits the sample into clubs based on the initial period (per capita) income levels of the regions, and (analytical) log-marginal likelihood expressions to calculate posterior model probabilities for models involving splits based on different initial period income levels of the sample of regions. Deriving the log-marginal likelihood used for model comparison purposes here involved a combined strategy that relied on: (i) analytical integration for some parameters of the model, (ii) numerical integration over the space and time dependence parameters, and (iii) fixing the variance ratio for the random effects versus noise vector.

Results of applying the model comparison procedure to a model that relied on dummy variable vectors to split the sample of 216 European regions according to initial period income levels were reported. They suggest strong evidence of two clubs or regimes based on regions whose 1995 level of per capita GVA was below and above 16,000.

Assuming two clubs, the second step of the approach involved estimating a space-time dynamic panel data model that used (logged) levels of regional income as the dependent variable and (logged) levels of previous period endowments of physical, knowledge and human capital stocks. Analytical expressions from Debassy, Ertur and LeSage (2011) for the partial derivatives showing dynamic response elasticities were used to examine the response of regional income levels over space and time to changes in initial period endowments. These dynamic responses provide clear evidence of the distinct long-term behaviour of the two clubs of regions.

Acknowledgements

The authors would like to acknowledge the Austrian Science Fund (FWF) for financial support (grant no. P19025-G11). We would like to thank Olivier Parent for his helpful comments on this manuscript, especially regarding the log-marginal likelihood derivation.

References

Abreu, M., de Groot, H.L.F. and R.J.G.M. Florax. 2005. Space and growth: A survey of empirical evidence and methods. Région et Dévelopment 21(1): 13-44.

Anselin, L. 2001. Spatial econometrics. In: Baltagi, Badi H. (Ed.), A companion to theoretical econometrics. Blackwell Publishers Ltd, Massachusetts, pp. 310-330.

Azariadis, C. and A. Drazen. 1990. Threshold externalities in economic development. The Quarterly Journal of Economics 105(2): 501-526.

Barro, R. 1991. Economic growth in a cross section of countries. The Quarterly Journal of Economics 106(2): 407-443.

Baumol, W.J. 1986. Productivity growth, convergence and welfare: What the long run data show? American Economic Review 76(5): 1072-1084.

Caballero J.R. and J.B. Jaffe. 1993. How high are the giants shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. In: U. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual 1993, Vol. 8.*. Cambridge [MA]: The MIT Press, pp. 15-74.

Canova, F. 2004. Testing for convergence clubs in income per capita: A predictive density approach. International Economic Review 45(1): 49-77.

Corrado, L. Martin, R. and M. Weeks. 2005. Identifying and interpreting regional convergence clusters across Europe. The Economic Journal 115(502): C133-C160.

Dall'erba, S. 2005. Productivity convergence and spatial dependence among Spanish regions. Journal of Geographical Systems 7(2): 207-227.

De Siano, R. and M. D'Uva. 2006. Club convergence in European regions. Applied Economics Letters 13(9): 569-574.

Debarsy, N., C. Ertur and J. P. LeSage. 2012. Interpreting dynamic space-time panel data models, Statistical Methodology 9(1-2): 158-171.

Desdoigts, A. 1999. Patterns of economic development and the formation of clubs. Journal of Economic Growth 4(3): 305-330.

Durlauf, S. 1996. A theory of persistent income inequality. Journal of Economic Growth 1(1): 75-94.

Durlauf, S. and P.A. Johnson 1995. Multiple regimes and cross-country growth behaviour. Journal of Applied Econometrics 10(4): 365-384.

Durlauf, S., Johnson, P.A. and J.R.W. Temple. 2005. Growth economics. In: P. Aghion and S.N. Durlauf (eds.), Handbook of Economic Growth, Volume IA. Amsterdam: Elsevier, pp. 555-677.

Ertur, C. and W. Koch. 2007. Growth, technological interdependence and spatial externalities: Theory and evidence. Journal of Applied Econometrics 22(6): 1033-1062.

Ertur, C., J. Le Gallo and C. Baumont. 2006. The European regional convergence process, 1980-1995: do spatial dependence and spatial heterogeneity matter? International Regional Science Review 29(1): 2-34.

Fingleton, B., and E. López-Bazo. 2006. Empirical growth models with spatial effects. Papers in Regional Science 85(1): 177-198.

Fischer M.M. 2011. A spatial Mankiw-Romer-Weil model: Theory and evidence. Annals of Regional Science 47(2): 419-436.

Fischer, M.M. and C. Stirböck. 2006. Pan-European regional income growth and club-convergence. Insights from a spatial econometric perspective. Annals of Regional Science 40(4):693-721.

Fischer, M.M., T. Scherngell and E. Jansenberger. 2005. Patents, patent citations and the geography of knowledge spillovers in Europe. In: T. Markowski (ed.), *Regional Scientists' Tribute to Professor Ryszard Domanski.* Warsaw, Poland: Polish Academy of Sciences, Committee for Space Economy and Regional Planning, pp.57-75.

Galor, O. 1996. Convergence? Inference from theoretical models. The Economic Journal 106(37): 1056-1069.

Hobijn, B. and P.H. Franses. 2000. Asymptotically perfect and relative convergence of productivity. Journal of Applied Econometrics 15(1): 59-81.

Islam, N. 2003. What have we learnt from the convergence debate? Journal of Economic Surveys 17(3): 309-362.

Islam, N. 1998. Growth empirics: A panel data approach a reply. The Quarterly Journal of Economics 113(1): 325-329.

Islam, N. 1995. Growth empirics: A panel data approach. The Quarterly Journal of Economics 110(4): 1127-1170.

Jaffe, A. B., M. Trajtenberg and R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108(3): 577-98.

Lee, K, Pesaran, M. and R. Smith. 1997. Growth and convergence in a multi-country empirical stochastic Solow model. Journal of Applied Econometrics 12(4): 357-392.

LeSage, J.P. and M.M. Fischer. 2008. Spatial growth regressions: model specification, estimation and interpretation. Spatial Economic Analysis 3(3): 275-304.

LeSage, J. P. and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. London: Taylor & Francis/CRC Press.

LeSage, J. P. and O. Parent. 2007. Bayesian model averaging for spatial econometric models. Geographical Analysis 39(3): 241-267.

Magrini, S. 2004. Regional (di)convergence. In V. Henderson and J.F. Thisse (eds.), Handbook of Regional and Urban Economics, Amsterdam: Elsevier, pp. 2741-2796.

Mankiw, N.G., Romer, D. and D. Weil. 1992. A contribution to the empirics of economic growth. The Quarterly Journal of Economics 107(2): 407-437.

Mora, T. 2005. Evidencing European regional convergence clubs with optimal grouping criteria. Applied Economics Letters 12(15): 937-940.

Parent, O. and J. P. LeSage. 2011. A space-time filter for panel data models containing random effects. Computational Statistics & Data Analysis 55 (1): 475-490.

Parent, O. and J. P. LeSage. 2010. A spatial dynamic panel model with random effects applied to commuting times. Transportation Research Part B: Methodological 44B(5): 633-645.

Parent, O. and J. P. LeSage. 2009. Spatial dynamic panel data models with random effects. Forthcoming in Regional Science & Urban Economics. Available at SSRN: http//ssrn.com/abstract=1609748.

Parent, O. and J. P. LeSage. 2008. Using the variance structure of the conditional autoregressive specification to model knowledge spillovers. Journal of Applied Econometrics 23(2): 235-256.

Phillips, P.C.B. and D. Sul. 2009. Economic transition and growth. Journal of Applied Econometrics 24(7): 1153-1185.

Quah, D. 1996. Convergence empirics across countries with (some) capital mobility. Journal of Economic Growth 1(1): 95-124.

Rey, S.J. and J. Le Gallo. 2009. Spatial analysis of economic convergence. In: T.C. Mills and K. Patterson (eds.), *The Palgrave Handbook of Econometrics Volume II: Applied Econometrics*. New York, USA: Palgrave-MacMillan, pp. 1251-1292.

Wansbeek T.J. and A. Kapteyn. 1982. A simple way to obtain the spectral decomposition of variance components models for balanced data. Communications in Statistics Part A - Theory Methods 11(18): 2105-2112.

Yu, J., R. de Jong and L.F. Lee. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146: 118-134.

Yu, J. and L.F. Lee. 2012. Convergence: A spatial dynamic panel data approach. Global Journal of Economics 1(1): 1250006-12500042.

Zellner, A. 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons.

Zellner, A. 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: P. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. North-Holland/Elsevier: Amsterdam, pp. 233-243.

Appendix A

Deriving the log-marginal likelihood used for model comparison purposes in our study involves a combined strategy that relies on analytical integration for some parameters of the model, numerical integration over the space and time dependence parameters, and fixing the variance ratio for the random effects versus noise vector. We will develop the log-marginal likelihood expressions to calculate posterior probabilities for models involving splits based on different initial period income levels of the sample of regions. Let us start with the task of analytically integrating out the parameters $\delta = (\ \psi \quad \tilde{\psi} \quad \alpha \quad \tilde{\alpha} \quad \beta \quad \tilde{\beta}\ )'$.

Proceeding to the task of analytically integrating out the parameters δ, we can concentrate out the parameters δ using:

$$\hat{\delta} \;=\; (Z'Z)^{-1}Z'Pg$$

which can be strategically written using the following expressions:

$$\begin{aligned}
\hat{\delta} &= (\delta_0 - \phi\delta_\phi - \rho\delta_\rho - \theta\delta_\theta) \\
\delta_0 &= (Z'Z)^{-1}Z'(F \otimes I_N)g \\
\delta_\phi &= (Z'Z)^{-1}Z'(L \otimes I_N)g \\
\delta_\rho &= (Z'Z)^{-1}Z'(F \otimes W)g \\
\delta_\theta &= (Z'Z)^{-1}Z'(L \otimes W)g
\end{aligned}$$

where

$$L \;=\; \begin{pmatrix} -1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -1 & 0 \end{pmatrix}$$

$$F = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & & 0 & 1 \end{pmatrix}$$

with L and F being $(T-1) \times T$ matrices.

Now consider the errors: $e = Pg - Z\delta$, which can be written using:

$$\begin{aligned}
e &= \begin{pmatrix} 1 & -\phi & -\rho & -\theta \end{pmatrix} \begin{pmatrix} E^{(1)} \\ E^{(2)} \\ E^{(3)} \\ E^{(4)} \end{pmatrix} \\
E^{(1)} &= (F \otimes I_N)g - Z(Z'Z)^{-1}Z'(F \otimes I_N)g \\
E^{(2)} &= (L \otimes I_N)g - Z(Z'Z)^{-1}Z'(L \otimes I_N)g \\
E^{(3)} &= (F \otimes W)g - Z(Z'Z)^{-1}Z'(F \otimes W)g \\
E^{(4)} &= (L \otimes W)g - Z(Z'Z)^{-1}Z'(L \otimes W)g \\
e'\Omega^{-1}e &= \tau' Q \tau \\
\tau &= \begin{pmatrix} 1 & -\phi & -\rho & -\theta \end{pmatrix} \\
Q_{ij} &= \mathrm{tr}(E^{(i)\prime}\Omega^{-1}(\lambda)E^{(j)}), \quad i = 1, \dots, 4 \quad j = 1, \dots, 4
\end{aligned}$$

The advantage of this specification is that the likelihood can be written expressing the sum of squared residuals $e'\Omega^{-1}e$ as a function of only the parameters ϕ, ρ, θ in the vector τ and the parameter λ, plus sample data information g, Z, W.

We assign an inverse gamma prior $IG(a, b)$ for σ_μ^2/λ:

$$\pi_s(\sigma_\mu^2/\lambda) \sim \frac{(ab/2)^{a/2}}{\Gamma(a/2)} (\sigma_\mu^2/\lambda)^{-(\frac{a+2}{2})} \exp(-\frac{ab}{2\sigma_\mu^2/\lambda}),$$

where a, b are parameters of the inverse gamma prior. We follow LeSage and Parent (2007) and assign Zellner's g-prior (Zellner 1986) to the parameters δ:[15]

$$\begin{aligned}
\pi_d(\delta|\sigma_\mu^2/\lambda) &\sim N(0, (\sigma_\mu^2/\lambda)V^{-1}) \\
V &= GZ'Z
\end{aligned}$$

Using Bayes theorem the marginal likelihood for the model can be written as the integral below which is analogous to that from LeSage and Parent (2007), where we use $\mathcal{D}$ to denote the data g, Z, W.

$$\begin{aligned}
& \int \pi_d(\delta|\sigma_\mu^2/\lambda)\pi_s(\sigma_\mu^2/\lambda)p(\mathcal{D}|\alpha,\delta,\rho,\phi,\theta,\sigma_\mu^2/\lambda)\ d\delta\ d\sigma_\mu^2/\lambda)\ d\rho\ d\phi\ d\theta \\
= \ & \kappa(2\pi)^{-(N(T-1)+K)/2}((T-1)\lambda+1)^{-N}|V|^{1/2} \\
\times \ & \int |I_N - \rho W|^T \frac{\lambda^{[N(T-1)]+a+2K+1}}{\sigma_\mu^2} \\
\times \ & \exp\left(-\frac{1}{2\sigma_\mu^2/\lambda}[ab + e'\Omega^{-1}e + \delta' V\delta + (\delta - \hat{\delta}(\phi,\rho,\theta))'(Z'Z)(\delta - \hat{\delta}(\phi,\rho,\theta))]\right) \\
\times \ & \pi_d\pi_\phi\pi_\rho\pi_\theta d\delta\ d\phi\ d\rho\ d\theta \\
\kappa = \ & \Gamma\left(\frac{a}{2}\right)^{-1}\left(\frac{ab}{2}\right)^{a/2}
\end{aligned}$$

Following LeSage and Parent (2007) we can use the properties of the multivariate normal pdf and the inverted gamma pdf to analytically integrate out the parameters δ and σ_μ^2/λ which produces an expression for the marginal likelihood as a function of the three parameters $\zeta = (\phi, \rho, \theta)$ only.

An expression that is analogous to that from LeSage and Parent (2007) arises:

$$\begin{aligned}
p(\zeta|\mathcal{D}) &= \tilde{\kappa}(\frac{G}{1+G})^{K/2}(T\lambda+1)^{-N} \\
&\times \int |I_N - \rho W|^T [ab + R(\zeta) + S(\zeta)]^{-[N(T-1)+a-1]/2}\pi_\phi\pi_\rho\pi_\theta\ d\phi\ d\rho\ d\theta
\end{aligned}$$

[15]See LeSage and Parent (2007) for the motivation for this type of prior.

where

$$
\begin{aligned}
\tilde{\kappa} &= \frac{\Gamma[(N(T-1)+a-1)/2]}{\Gamma(a/2)}(ab)^{a/2}\pi^{-[N(T-1)-1]/2} \\
R(\zeta)+S(\zeta) &= \frac{1}{G+1}\tau' Q\tau \\
&+ \frac{G}{G+1}\left(Pg-\hat{\alpha}\iota_{NT}\right)'\left(Pg-\hat{\alpha}\iota_{NT}\right) \\
\hat{\alpha} &= U^{(1)}-\phi U^{(2)}-\rho U^{(3)}-\theta U^{(4)} \\
U^{(1)} &= (F\otimes I_N)y \\
U^{(2)} &= (L\otimes I_N)y \\
U^{(3)} &= (F\otimes W)y \\
U^{(4)} &= (L\otimes W)y
\end{aligned}
$$

with Γ denoting the gamma function. Recall that $e'\Omega^{-1}e = \tau' Q\tau$ and Ω is a function of λ which we are treating as a fixed scalar, so Ω is presumed known. Without loss of generality we can view λ as equal to any fixed value here, but in practice we should test for robustness across various values of this parameter reflecting the variance ratio of the random effects to noise.

While we developed these expressions for the case of unrestricted θ, we can reduce the trivariate numerical integration problem to a bivariate problem by imposing the restriction $\theta = -\rho\phi$, which is the approach we take in our application.

Appendix B

We illustrate the model comparison procedure here using a generated vector of growth rates constructed using sample data from our 216 EU regions. The growth rates relationship in (6) was estimated based on a dummy variable vector splitting the sample at initial period income levels of $m = 20,000$. The parameter estimates were then used to produce predicted values that reflected two regimes with regions split at this income level. When generating predicted values, parameters $\hat{\rho} = 0.65, \hat{\phi} = -0.18$ were used,in conjunction with a value of $\theta = 0.025$, which does not obey the restriction on the parameter $\theta = -\phi\rho$. Specifically,

$\theta = -\phi\rho = -(-0.18 \cdot 0.65) = 0.117$, rather than the value $\theta = 0.025$ used to produce a sample of growth rates. However, posterior model probabilities were calculated based on the assumption that $\theta = -\phi\rho$, as a test of the impact on performance in this type of setting where the assumption is violated.

The distributions of growth rates for the two clubs that resulted from this approach are shown in Figure 3, where we see the high income club exhibiting a slightly lower mean growth rate than the lower income club. This is of course consistent with the usual notion of β-convergence, where regions with lower initial levels of income exhibit higher growth rates than higher income regions.

The estimated ratio of variances $\hat{\lambda} = \hat{\sigma}^2_\mu / \hat{\sigma}^2_\varepsilon$ equalled 0.2594. Table 9 shows posterior model probabilities derived from a comparison of models based on splits of the regions ranging from $10,000$ to $32,000$ in increments of $2,000$, for values of $\hat{\lambda}$ as well as $(1/2)\ \hat{\lambda}$ and $2\hat{\lambda}$.

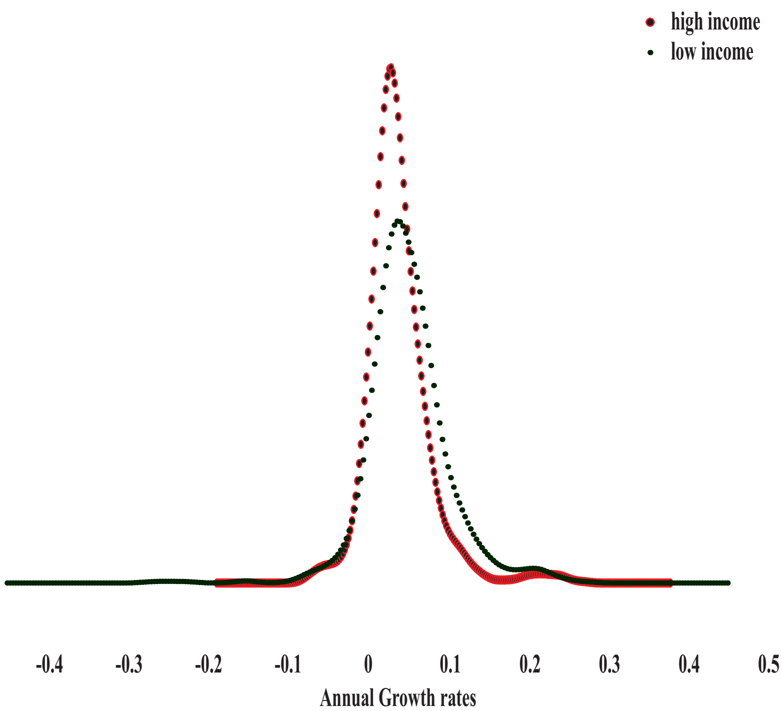

Figure 3: Distribution of growth rates generated using $m = 20,000$

The resulting posterior model probabilities point to the correct model based on a split of the regions at the $m = 20,000$ level for all three settings of λ. As we would expect, there is some degradation of performance for values based on $1/2\lambda$ and 2λ, but the correct inference would be drawn in these cases.

Table 9: Posterior model probabilities for generated data example

Sample split y_0 levels Model $= q$	Prob(model $= q$) $\lambda = (1/2)\hat{\lambda}$	Prob(model $= q$) $\lambda = \hat{\lambda}$	Prob(model $= q$) $\lambda = 2\hat{\lambda}$
10,000	0.0000	0.0000	0.0000
12,000	0.0000	0.0000	0.0000
14,000	0.0001	0.0000	0.0000
16,000	0.0000	0.0000	0.0000
18,000	0.2744	0.1157	0.2495
20,000*	0.6723	0.8658	0.6689
22,000	0.0009	0.0010	0.0005
24,000	0.0523	0.0000	0.0000
26,000	0.0000	0.0000	0.0000
28,000	0.0000	0.0000	0.0811
30,000	0.0000	0.0175	0.0000
32,000	0.0000	0.0000	0.0000
* indicates model that generated the growth rates			